多群判別問題に対する新解法(モデリングと最適化の理論)

北原, 知就; 水野, 眞治; 中田, 和秀

数理解析研究所講究録
1526巻 2006年 120-128

多群判別問題に対する新解法*

北原 知就†， 水野 眞治‡， 中田 和秀§

1 はじめに

最近，Lanckriet et al. [7] によって2クラス線形判別問題に対するミニマックスアプローチが開発された．それにおいて必要となる入力データは各クラスの平均ベクトル，分散共分散行列のみである．一般に，判別関数の精度は，各クラスの平均ベクトル，分散共分散行列だけでなく，各クラスがどのような分布形を取るかに影響される．このことは，場合によっては判別関数の精度が非常に悪くなってしまう可能性があることを意味する．そこで，Lanckriet et al. は与えられた平均ベクトル，分散共分散行列のもとでの最悪の (マックス) 場合の誤判別率が最小に (ミニ) なるように線形の判別関数を決定することを提案した．彼らのアルゴリズムはミニマックス確率マシン (Minimax Probability Machine) と呼ばれている．Lanckriet et al. はこのような判別関数を求める問題が二次錐計画問題に帰着され，効率的に解を得ることができることを示した．さらに，計算された判別関数が実際の問題に対して有効に機能することを実証した．このような理論的，実用的な側面から，最近 MPM は注目を集め，活発な研究がなされている [4][5].

本稿は，Lanckriet at al. のミニマックスアプローチの，クラスの数が3つ以上の多クラスへの拡張について述べる．他の2クラス判別手法がその多クラスへの拡張が良く研究されているのに対し，これに関してはほとんど研究されていない．その自然な拡張としては，多クラスにおいても判別関数の最悪の場合の誤判別率を最小にするように判別関数を決定する，ということが考えられるが，これを実際に求めることは困難である．そこで本稿では最悪の場合のペアごとの誤判別率という新たな指標を導入した．これは最悪の場合の誤判別率の下界であり，ある条件の下で最悪の場合の誤判別率をよく近似する，という性質を持つ．このような性質に注目し，本稿では最悪の場合の誤判別率ではなく，最悪の場合のペアごとの誤判別率を最小化するように判別関数を決定することを提案する．このような判別関数を決定する問題はパラメータ付き二次錐計画問題に帰着され，効率的に解くことができる．さらに，ベンチマーク問題における予備数値実験によりサポートベクターマシンに匹敵する判別率が得られることを示す.

本稿の構成は以下のとおりである．まず次節で Lanckriet et al. [7] を参考に彼らのミニマックスアプローチについて簡単に説明し，続く3節でその多クラスへの拡張について説明する．4節では問題がパラメータ付き二次錐計画問題に帰着されることを示す．得られた問題はその特殊な性質に注目することで容易に解くことができる．このことについて5節で述べる．6節で予備数値実験の結果を説明し，最後に7節でまとめを行う.

本稿では n 次元実ベクトル空間を $\Re^n$ と表し，n 次元実対称正定値行列の空間を $\mathbb{S}^n_{++}$ と表す.

---

*本稿は Kitahara et al. [6] の内容を再構成し，まとめたものである.

†東京工業大学　社会理工学研究科　経営工学専攻
‡東京工業大学　社会理工学研究科　経営工学専攻
§東京工業大学　社会理工学研究科　経営工学専攻

2 ミニマックス確率マシン

2つのクラス Class 1と Class 2があり，そのそれぞれの平均ベクトル，分散共分散行列を $(\mu_i, \Sigma_i) \in \Re^n \times \mathrm{S}^n_{++}$ とする．本稿では，分散共分散行列の正定値性を仮定する．実際上は，分散共分散行列に標準化項を加えるのが普通であるので，この仮定は成立する．その他の分布に関する仮定は置かない．

いま，判別関数 $f(z) = a^T z + b$ $(a \in \Re^n \setminus \{0\}, b \in \Re, z \in \Re^n)$ を定めて，新たに得られたサンプル $x \in \Re^n$ を以下のように判別する．すなわち，$f(x) < 0$ であれば Class 1に，$f(x) > 0$ であれば Class 2にそれぞれ判別し，$f(x) = 0$ となった場合はどちらに判別しても良いとする．

判別関数を決定する際, なるべく判別精度を高くしたいというのは自然な考え方である．しかし, 各クラスの平均ベクトルと分散共分散行列しかわかっていない場合, 判別の精度は各クラスがどのような分布形を取るかによって変わってくる．従って, ある分布形のもとで高い判別精度を示す判別関数が，別の分布形のもとでは非常に低い精度を示すということが起こりうる．このような分布形による判別精度の不確定性に対し，Lanckriet et al. [7] のミニマックスアプローチはさまざまに分布形を動かしたときの最悪の場合の (マックス) 誤判別率が最小 (ミニ) になるような判別関数を採用することを提案する．判別関数 $f(z) = a^T z + b$ に対して，Class 1のサンプルを Class 2であると誤って判別する最悪の場合の確率は,

$$\sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{a^T x + b \geq 0\},$$

と表される．ここで，$\sup_{x \sim (\mu_1, \Sigma_1)}$ という表記は，平均ベクトルが μ_1，分散共分散行列が Σ_1 であるようなすべての確率分布に対して上限をとることを意味する．もう一方の場合も同様に考えると，判別関数 $a^T z + b$ の最悪の場合の誤判別率 α は,

$$\alpha = \max\{ \sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{a^T x + b \geq 0\}, \sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{a^T x + b \leq 0\}\},$$

のように与えられる．上述したミニマックスアプローチは,

$$\min \ \alpha,$$

と書くことができ，さらに上で導入した表記を用いれば,

$$\begin{array}{ll} \min & \alpha \\ \text{subject to} & \sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{a^T x + b \geq 0\} \leq \alpha, \\ & \sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{a^T x + b \leq 0\} \leq \alpha, \end{array} \tag{1}$$

と表すことができる．この問題における変数は $\alpha \in [0,1]$，$a \in \Re^n$，$b \in \Re$ である．Lanckriet et al. [7] は $\mu_1 \neq \mu_2$ であれば，問題 (1) の任意の最適解 (α, a, b) において $\alpha < 1$ かつ $a \neq 0$ であることを示した．問題 (1) における制約条件は一見したところ厄介であるが，次に述べる重要な命題を利用することで簡単に扱うことができる．

命題 1 *(Lanckriet et al. [7])* いま，$(\mu, \Sigma) \in \Re^n \times \mathrm{S}^n_{++}$ が与えられているとする．このとき，$a \neq 0$ であるような任意の $(a, b) \in \Re^n \times \Re$ に対して,

$$\sup_{x \sim (\mu, \Sigma)} \Pr\{a^T x + b \geq 0\} = \frac{a^T \Sigma a}{a^T \Sigma a + s^2},$$

が成立する．ここで，$s \equiv \max\{-b - a^T \mu, 0\}$ である．

命題 1 を利用すると問題 (1) は次の問題に帰着することができる (詳細は [7] を参照)：

$$\begin{array}{ll} \min & \|\Sigma_1^{1/2} a\| + \|\Sigma_2^{1/2} a\| \\ \text{subject to} & (\mu_2 - \mu_1)^T a = 1. \end{array} \tag{2}$$

この問題は二次錐計画問題であり，内点法を使って効率的に解くことができる．このようにして判別関数を得るアルゴリズムはミニマックス確率マシン (Minimax Probability Machine, MPM) と呼ばれる．Lanckriet et al. [7] では，MPM が実際の問題に対し有効に機能すること，およびそれがサポートベクターマシン (Support Vector Machine, SVM) に匹敵する判別精度を示す，という数値実験の結果が報告されている．

3 ミニマックスアプローチの多クラスへの拡張

現実的には，クラスの数は必ずしも 2 つとは限らない．本節では前節で概説したミニマックスアプローチの多クラスへの拡張を説明する．まず 3.1 節で多クラス判別における判別関数の置き方について述べる．続く 3.2 節では，まず最悪の場合の誤判別率を最小にするような線形判別関数を求める問題を定義する．これは前節で述べたミニマックスアプローチの自然な拡張であるが，これを解くことは困難である．そこで 3.3 節では最悪の場合の誤判別率の近似値である最悪の場合のペアごとの誤判別率を導入し，これを最小にする問題を提示する．

3.1 多クラス判別における判別関数

$m \geq 2$ に対して添え字集合 $M = \{1, 2, \ldots, m\}$ を定義する．いま，m 個のクラスがあるとし，$i \in M$ に対して第 i 番目のクラスを Class i と呼ぶ．ここで，各 i に対して平均ベクトル $\mu_i \in \Re^n$ および分散共分散行列 $\Sigma_i \in \mathrm{S}^n_{++}$ が既知であるとする．2 クラスの場合同様，各クラスの分布形に対しては何も仮定しない．

前節で述べた 2 クラス判別の場合，一つの判別関数 $f(z) = a^T z + b$ を用い，新たに得られたサンプル $x \in \Re^n$ を $f(x)$ の符号によって判別した．いま，$f(z) = f_1(z) - f_2(z)$ を満たす適当な関数を考えれば，条件 $f(z) < 0$ は $f_1(z) < f_2(z)$ と等価である．したがって，上のような判別規則はサンプル x を $f_1(x) < f_2(x)$ ならば Class 1 に，$f_1(x) > f_2(x)$ ならば Class 2 に判別することと同じである．このように考えると 2 クラスの場合の判別規則を多クラスに自然に拡張することができる．すなわち，多クラスの場合，すべての $i \in M$ に対して線形の判別関数

$$f_i(z) = a_i^T z + b_i, \tag{3}$$

を考える．ここで $a_i \in \Re^n$，$b_i \in \Re$ である．これらによって，新たに得られたサンプル $x \in \Re^n$ を判別関数の値が最も小さいクラスへと判別する．形式的に表すと，サンプル x が Class l に判別されるとき，次の関係式を満たしている：

$$f_l(x) = \min_{i \in M} f_i(x).$$

上の式を満たす添え字が複数ある場合には，そのうちのどのクラスに判別しても良いものとする．本稿ではこのような判別規則を採用する．問題はどのように m 個の線形判別関数 $f_i(z) = a_i^T z + b_i$，$i \in M$ を決定するか，ということである．

3.2 最悪の場合の誤判別率の最小化問題

各 $i \in M$ に対して，次の集合

$$R_i \equiv \{z \in \Re^n | a_i^T z + b_i < a_j^T z + b_j,\ \forall j \in M, j \neq i\},$$

を定義する．明らかに，任意のサンプル x は R_i (またはその閉包) のうちの一つに属する．本稿で採用する規則に従えば，サンプル x は $x \in R_i$ を満たすとき Class i に判別される．前節で導入した表記を用いると，任意の i に対して，Class i のサンプルを誤って判別する最悪の場合の確率 (以下，Class i の最悪の場合の誤判別率と呼ぶ) は

$$\alpha_i = \sup_{x \sim (\mu_i, \Sigma_i)} \Pr\{x \notin R_i\}, \tag{4}$$

と表される．判別関数の最悪の場合の誤判別率は

$$\alpha = \max_{i \in M} \alpha_i, \tag{5}$$

と書くことができ，これを最小化する問題は

$$\begin{array}{ll} \min & \alpha \\ \text{subject to} & \sup_{x \sim (\mu_i, \Sigma_i)} \Pr\{x \notin R_i\} \leq \alpha, \end{array} \tag{6}$$

と表すことができる．ここで，変数は $\alpha \in [0,1]$ および $(a_i, b_i) \in \Re^n \times \Re$ $(i \in M)$ である．このように，多群においても最悪の場合の誤判別率を最小にするような判別関数を求めようと考えるのは Lanckriet et al. のアプローチの多群への自然な拡張と考えられる．しかしながら問題 (6) を解くのは容易ではなく，これまでのところこれを解く方法は得られていない．

3.3 最悪の場合のペアごとの誤判別率

Class i の最悪の場合の誤判別率 α_i の代わりに，次式で定義される Class i の最悪の場合のペアごとの誤判別率を導入する:

$$\beta_i \equiv \max_{j \neq i} \sup_{x \sim (\mu_i, \Sigma_i)} \Pr\{a_i^T x + b_i \geq a_j^T x + b_j\}. \tag{7}$$

次の命題は β_i が α_i の下界であることを示している．

命題 2 *(Kitahara et al. [6])* それぞれの i に対して *Class* i の最悪の場合の誤判別率と最悪の場合のペアごとの誤判別率をそれぞれ式 *(4)* と式 *(7)* で定義するとき，次の関係が成立する：

$$\beta_i \leq \alpha_i \leq (m-1)\beta_i \tag{8}$$

上の命題より，クラスの数 m か Class i の最悪の場合の誤判別率 α_i が小さい場合，Class i のペアごとの最悪の場合の誤判別率 β_i は α_i の良い近似となっていることがわかる．特に，$m = 2$ のとき，β_i は α_i に一致する．そこで，最悪の場合のペアごとの誤判別率

$$\beta \equiv \max_{i \in M} \beta_i \tag{9}$$

を最小にするような判別関数を求めることを考える．そのような判別関数は次の問題

$$
\begin{array}{ll}
\min & \beta \\
\text{subject to} & \sup_{x\sim(\mu_i,\Sigma_i)}\{a_i^T x + b_i \geq a_j^T x + b_j\} \leq \beta,\ i,j \in M\ j \neq i
\end{array}
\tag{10}
$$

を解くことによって得られる．この問題における変数は $\beta \in [0,1]$ および $(a_i, b_i) \in \Re^n \times \Re$ $(i \in M)$ である．この問題に関して，次の定理が成立する．

定理 3 *(Kitahara et al. [6]*[1] *)* 平均ベクトルが等しい *2* つのクラスが存在すれば *(*すなわち，ある $i,j \in M$ $i \neq j$ が存在して $\mu_i = \mu_j$ となる*)*，問題 *(10)* の最適値は *1* である．そうでなければ，問題 *(10)* の任意の最適解において，$\beta < 1$ およびすべての $i \in M$ に対して $\mu_i \in R_i$ が成立する．

平均ベクトルが等しくなるようなクラスが存在する場合，問題 (10) は意味のある解を持たないので，これ以後，各クラスの平均ベクトルはすべて異なるものと仮定する．

4 パラメータ付き二次錐計画問題への帰着

この節では，問題 (10) がパラメータ付き二次錐計画問題に帰着されることを示す．

定理 3 から，問題 (10) に制約 $\mu_i \in R_i$ $(i \in M)$，すなわち，任意の $i,j \in M$, $i \neq j$ について

$$a_i^T \mu_i + b_i < a_j^T \mu_i + b_j, \tag{11}$$

を加えることができる．ここで，次の二つの制約

$$a_i^T \mu_i + b_i < a_j^T \mu_i + b_j$$

および

$$a_j^T \mu_j + b_j < a_i^T \mu_j + b_i$$

から

$$(a_j - a_i)^T(\mu_i - \mu_j) > 0,$$

となることがわかる．これは任意の相異なる任意の $i,j \in M$ に対して $a_i \neq a_j$ が成立することを意味する．したがって，命題 1，および任意の相異なる $i,j \in M$ に対して $-(a_i - a_j)^T\mu_i - (b_i - b_j) > 0$ であるから，問題 (10) の制約は

$$-(a_i - a_j)^T\mu_i - (b_i - b_j) \geq \eta(\beta)\|\Sigma_i^{1/2}(a_i - a_j)\|,\ \eta(\beta) \equiv \sqrt{\frac{1-\beta}{\beta}},$$

と等価である．以上より，問題 (10) は次の問題に帰着される：

$$
\begin{array}{ll}
\min & \beta \\
\text{subject to} & -(a_i - a_j)^T\mu_i - (b_i - b_j) \geq \eta(\beta)\|\Sigma_i^{1/2}(a_i - a_j)\|,\ i,j \in M,\ j \neq i, \\
& -(a_i - a_j)^T\mu_i - (b_i - b_j) > 0,\ i,j \in M,\ j \neq i.
\end{array}
\tag{12}
$$

問題 (12) の制約は (a_i, b_i) $(i \in M)$ に関して正斉次である．すなわち，$\beta, (a_i, b_i)$ $(i \in M)$ が実行可能解であれば，任意の $s > 0$ に対して $\beta, (sa_i, sb_i)$ $(i \in M)$ もまた実行可能解となる．したがって，問題 (12) における制約

$$-(a_i - a_j)^T\mu_i - (b_i - b_j) > 0,\ i,j \in M,\ j \neq i,$$

[1] Kitahara et al. [6] では最悪の場合のペアごとの誤判別率ではなく 1 からこれを減じた値，すなわち最悪の場合のペアごとの正判別率を利用して議論を進めているが，証明は本質的に同じである．これは後出の命題 4 についても同様である．

を次の

$$-(a_i - a_j)^T \mu_i - (b_i - b_j) \geq 1,\ i,j \in M,\ j \neq i,$$

で置き換えることができる．以上から，次のパラメータ付き二次錐計画問題

$$\begin{array}{lll} \min & \beta & \\ \text{subject to} & -(a_i - a_j)^T \mu_i - (b_i - b_j) \geq \eta(\beta) \|\Sigma_i^{1/2}(a_i - a_j)\|,\ i,j \in M,\ j \neq i, & (13) \\ & -(a_i - a_j)^T \mu_i - (b_i - b_j) \geq 1,\ i,j \in M,\ j \neq i, & \end{array}$$

を得る．問題 (13) は非線形計画問題であり，一見解くのが難しいように思われるが，その特殊な性質に注目することで簡単に解くことができる．

5 パラメータ付き二次錐計画問題の解法

この節では，まず問題 (13) の 2 つの重要な性質を説明し，その後それに注目したアルゴリズムを説明する．

1 つ目の性質は，問題 (13) において β を固定すると制約は残りの変数に関する二次錐制約になる，というものである．ここで，変数 $x \in \Re^n$ に関する二次錐制約は一般に

$$c^T x + d \geq \|Ax + b\| \tag{14}$$

と表される．ここで $A \in \Re^{m \times n}$, $b \in \Re^m$, $c \in \Re^n$, $d \in \Re$ は与えられたデータである．制約 (14) を満たす解が存在するかどうかを調べるには，次の問題を解けばよい：

$$\begin{array}{ll} \min_{x,t} & t \\ \text{subject to} & c^T x + d + t \geq \|Ax + b\|, \\ & t \geq 0. \end{array} \tag{15}$$

ここで，$t \geq 0$ は新たな変数である．これは二次錐計画問題であり，内点法を利用して効率的に解くことができる．制約 (14) は問題 (15) の最適値が 0 のとき，そしてそのときに限り実行可能である．このとき，実行可能解は問題 (15) を解くことによって得られる．

2 番目の性質は，次の命題が示すように問題 (13) がある種の単調性を持つ，ということである．

命題 4 *(Kitahara et al. [6])* 問題 *(13)* がある $\beta' \in (0,1]$ で実行不能であれば，任意の $\beta \in (0,\beta']$ で実行不能である．

命題 4 で述べた性質によって，問題 (13) を二分法を利用して解くことができる．以下では，まずアルゴリズムの基本的なアイディアを述べ，その後形式的な表現を与える．

問題 (13) の最適値を β^* とする．始めは β^* は区間 $(0,1]$ に存在することしかわかっていない．まず，この区間の中央値である $\beta = 0.5$ をとり，この値で問題 (13) が実行可能であるかどうかを調べる．本稿ではこのような操作を **bisection**$((0,1])$ と表す．形式的には，区間 $(\beta_l, \beta_u] \subset (0,1]$ に対して，

$$\textbf{bisection}((\beta_l, \beta_u]) = \begin{cases} 1, & \text{問題 (13) が} \beta = \frac{\beta_l + \beta_u}{2} \text{に対して実行可能,} \\ 0, & \text{その他,} \end{cases}$$

と書くことができる．前述したように，**bisection**$((0,1])$ を行うには二次錐計画問題を解けばよい．もし **bisection**$((0,1]) = 1$ であれば，β^* は区間 $(0,0.5]$ に存在することがわかる．そうでない場合は，命題 4 から β^* は区間 $(0.5,1]$ に存在することが結論される．いずれの場合も，区間の幅を半

分にできることに注意する．次に，状況に応じて **bisection**$((0,0.5])$ または **bisection**$((0.5,1])$ を実行し，以後繰り返す．明らかに，以上のような操作の繰り返しで最適値の存在する区間の幅を任意に小さくすることができる．アルゴリズムの形式的な表現を表1にまとめた．

表 1: 2分探索アルゴリズム

```
入力：
平均ベクトル，分散共分散行列 (μ_i, Σ_i) (i = 1,...,m)；
精度パラメータ ε > 0；
初期区間 (β_l, β_u] = (0,1]；
初期解 (a_i, b_i) = (0,0) (i = 1....,m)；

while (β_u − β_l ≥ ε){
    if (bisection((β_l, β_u]) = 1){
        (β_l, β_u] ← (β_l, (β_l+β_u)/2]；
        update (a_i, b_i) (i = 1,...,m)；
    }
    else {
        (β_l, β_u] ← ((β_l+β_u)/2, β_u]；
    }
}

出力:
近似最適値 β_l；
解 (a_i, b_i) (i = 1,...,m)；
```

6　予備数値実験

本稿で提案したのは，多クラス判別問題において最悪の場合の振る舞いを考慮に入れた判別関数を構成する方法である．本節ではそのような提案手法がどの程度実用的かを明らかにするための予備実験について述べる．

本稿では平均ベクトルと分散共分散行列が既知であるという判別問題を扱ってきたが，ここではベンチマーク問題としてすべてのデータがわかっているものを使用した．各問題に対して，まずデータから平均ベクトルと分散共分散行列を推定した．これらの推定値をもとに判別関数を決定し，それを利用してデータの判別を行った．

ベンチマーク問題として，UCI Repository of machine learning databases [8] より **iris, wine, glass, vehicle**[2]を採用した．各問題の詳細を表2にまとめた．前節で述べたアルゴリズムにおける精度パラメータ ϵ は0.001とし，二次錐計画問題のソルバーは LOQO [9, 10] を利用した．

計算結果を表3にまとめた．表の第2列に示したのは1から最悪の場合のペアごとの誤判別率を減じた値である．これは，判別関数の最悪の場合のペアごとの正判別率を表す．第3列にまとめたのは，実際に判別を行ったときの精度である．表3より，各問題に対して，実際の判別率は $1-\beta$

[2]**vehicle** は UCI Repository より採用したが，もともとは Statlog collection によるものである．

表 2: ベンチマーク問題

問題	データ数	クラス数 (m)	説明変数の数 (n)
iris	150	3	4
wine	178	3	13
glass	214	6	9
vehicle	846	4	18

表 3: 計算結果

問題	$1-\beta$	判別精度	判別精度 (SVM，線形カーネル)
iris	0. 784	0. 973	0. 973
wine	0. 861	1. 000	0. 994
glass	0. 314	0. 645	0. 664
vehicle	0. 255	0. 766	0. 809

よりもかなり大きくなっていることがわかる．ここで，命題 2 から，α，β をそれぞれ判別関数の最悪の場合の誤判別率，最悪の場合のペアごとの誤判別率とすれば，$\beta \le \alpha$ が成立する．この不等式を辺ごとに 1 から減じれば，$1-\alpha \le 1-\beta$，すなわち，

(判別関数の最悪の場合の正判別率) $\le$ (判別関数の最悪の場合のペアごとの誤判別率)

という関係が成り立つ．これら 2 つの事実が示唆しているのは，判別関数の最悪の場合の正判別率と実際の判別率の間には大きな開きがあり，提案手法は実際の問題に対して非常によく機能する，ということである．

提案手法を他と比較するため表 3 の第 4 列に線形カーネルを用いたサポートベクターマシンによる判別結果を示した．この結果は [3] から引用した．提案手法の判別結果はサポートベクターマシンのそれに匹敵するものであることが見て取れる．サポートベクターマシンは有力な判別法の一つであり，このことは提案手法の有効性の証左といえる．

7 まとめ

平均ベクトルと分散共分散行列がわかっているいくつかのクラスを，判別関数を用いて判別するとき，その判別精度はそれぞれのクラスがどのような分布形を取るかに左右される．最近 Lanckriet et al. [7] によって提案されたミニマックスアプローチは，このような課題に対する一つの指針を与える．彼らのミニマックスアプローチは二つのクラスを線形関数で判別する際，その最悪の場合の (マックス) 誤判別率を最小に (ミニ) するように判別関数を決定するというものである．このような判別関数は関連する二次錐計画問題を解くことで簡単に得られる．判別関数を上記のように求めるアルゴリズムはミニマックス確率マシンと呼ばれ，現在，その理論的，実用的重要性から注目を集めている．

本稿ではこれまでほとんど研究されていなかった Lanckriet et al. のミニマックスアプローチの多クラスへの拡張について述べた．その自然な拡張としては，多クラスにおいても判別関数の最悪の場合の誤判別率 α を最小にするように判別関数を決定する，ということが考えられるが，これ

を実際に求めることは困難である．そこで本稿では α の近似値として最悪の場合のペアごとの誤判別率 β という新たな指標を導入し，β を最小にするように判別関数を決定することを提案した．この問題はパラメータ付き二次錐計画問題に帰着される．得られたパラメータ付き二次錐計画問題は特殊な性質を持ち，これに注目することで2分法で簡単に解くことができる．実施した予備数値実験の結果，提案手法がベンチマーク問題に対して有効に機能し，サポートベクターマシンに匹敵する判別率を示すことが確認された．

今後の課題としては，不正確な平均ベクトル，分散共分散行列を扱えるように提案手法のロバスト版を考えること，非線形の判別領域を許すように提案手法をカーネル化すること，などが挙げられる．

参考文献

[1] S. Boyd and L. Vandenberghe: *Convex Optimization* (Cambridge University Press, 2004).

[2] C-H. Hoi and M. R. Lyu: Robust face recognition using minimax probability machine. *Proceedings of the 2004 IEEE International Conference on Multimedia and Expo (ICME 2004)*, **2**, (2004), 1175–1178.

[3] C. W. Hsu and C. J. Lin: A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, **13**, (2002), 415–425.

[4] K. Huang, H. Yang, I. King, and M. R. Lyu: Learning Classifiers from Imbalanced Data Based on Biased Minimax Probability Machine. *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, **2**, (2004), 558–563.

[5] K. Huang, H. Yang, I. King, M. R. Lyu, and L. Chan: The Minimum Error Minimax Probability Machine. *Journal of Machine Learning Research*, **5** (2004), 1253–1286.

[6] T. Kitahara, S. Mizuno, and K. Nakata: An Extension of a Minimax Approach to Multiple Classification. *Technical Report*, Tokyo Institute of Technology, (2006). (Available from Optimization Online : http://www.optimization-online.org/)

[7] G. R. G. Lanckriet, L. El Ghaoui, C. Bhattacharyya, and M. I. Jordan: A Robust Minimax Approach to Classification. *Journal of Machine Learning Research*, **3** (2002), 555–582.

[8] D. J. Newman, S. Hettich, C. L. Blake, and C. J. Merz: UCI Repository of machine learning databases. Department of Information and Computer Sciences, University of California, Irvine, CA, 1998. http://www.ics.uci.edu/~mlearn/MLRepository.html

[9] R. J. Vanderbei: LOQO User's Manual –Version 4.05. *Technical Report ORFE 99*, Operations Research and Financial Engineering, Princeton University, (2000).

[10] R. J. Vanderbei and H. Yurttan: Using LOQO to solve second-order cone programming problems. *Technical Report SOR 98-9*, Statistics and Operations Research, Princeton University, (1998).